Atty. Dkt. No. 016778-0470

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Masataka ANDOH, et al.

Title: cDNA MICROARRAY DATA CORRECTION SYSTEM, METHOD, PROGRAM, AND MEMORY MEDIUM

Appl. No.: 10/696,572

Filing Date: 10/30/2003

Examiner: Unknown

Art Unit: 1651

CLAIM FOR CONVENTION PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

The benefit of the filing date of the following prior foreign application filed in the following foreign country is hereby requested, and the right of priority provided in 35 U.S.C. § 119 is hereby claimed.

In support of this claim, filed herewith is a certified copy of said original foreign application:

Japanese Patent Application No. 2003-124585
filed 04/28/2003.

Respectfully submitted,

Date: July 1, 2004

By Phillip J. Articola for / Reg. No. 38,819

David A. Blumenthal
Attorney for Applicant
Registration No. 26,257

FOLEY & LARDNER LLP
Customer Number: 22428
Telephone: (202) 672-5407
Facsimile: (202) 672-5399

US

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: ２００３年　４月２８日

出願番号
Application Number: 特願２００３－１２４５８５

[ST. 10/C]: [ＪＰ２００３－１２４５８５]

出願人
Applicant(s): 日本電気株式会社
大瀧　慈
社団法人バイオ産業情報化コンソーシアム

２００３年　９月２５日

特許庁長官
Commissioner,
Japan Patent Office 今井康夫

出証番号　出証特２００３－３０７８９６８

【書類名】　特許願
【整理番号】　64002125
【特記事項】　特許法第３０条第１項の規定の適用を受けようとする特許出願
【提出日】　平成15年 4月28日
【あて先】　特許庁長官殿
【国際特許分類】　G06N　7/00
【発明者】
【住所又は居所】　東京都港区芝五丁目７番１号　日本電気株式会社内
【氏名】　安東　正貴
【発明者】
【住所又は居所】　東京都港区芝五丁目７番１号　日本電気株式会社内
【氏名】　斎藤　彰
【発明者】
【住所又は居所】　広島県廿日市市宮園９丁目１の７
【氏名】　大瀧　慈
【発明者】
【住所又は居所】　広島県広島市佐伯区楽々園５－９　五日市住宅１６－２０２
【氏名】　佐藤　健一
【発明者】
【住所又は居所】　広島県広島市南区仁保南２丁目８－７
【氏名】　西山　正彦
【発明者】
【住所又は居所】　東京都中央区八丁堀二丁目２６番９号　グランデビルディング　社団法人バイオ産業情報化コンソーシアム内
【氏名】　大谷　敬子

【特許出願人】
　【識別番号】　000004237
　【氏名又は名称】　日本電気株式会社
【特許出願人】
　【識別番号】　503077165
　【氏名又は名称】　大瀧　慈
【特許出願人】
　【識別番号】　500535301
　【氏名又は名称】　社団法人バイオ産業情報化コンソーシアム
【代理人】
　【識別番号】　100071272
　【弁理士】
　【氏名又は名称】　後藤　洋介
【選任した代理人】
　【識別番号】　100077838
　【弁理士】
　【氏名又は名称】　池田　憲保
【手数料の表示】
　【予納台帳番号】　012416
　【納付金額】　21,000円
【提出物件の目録】
　【物件名】　明細書　1
　【物件名】　図面　1
　【物件名】　要約書　1
　【包括委任状番号】　0018587
【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　ｃＤＮＡマイクロアレイデータの補正システム、方法、プログラム、及び記録媒体

【特許請求の範囲】

【請求項１】　マイクロアレイデータのグローバル及びローカルな歪みに対してより精密な補正をし、さらに蛍光色素の感度の違いによる測定誤差を補正するｃＤＮＡマイクロアレイデータの補正システムにおいて、

各スポットのバックグラウンドノイズの除去及び信頼性を示すフラッグ情報を考慮し、あらかじめ調整されている遺伝子発現強度データを入力する入力装置と、　前記入力された遺伝子発現強度データに対して、グリッド毎の順序統計量を用いて遺伝子発現強度データを規準化し、規準化された規準化遺伝子発現強度データを送出するデータ規準化手段と、前記規準化遺伝子発現強度データに対して、グリッドの座標におけるスポット位置に依存する歪みをノンパラメトリック平滑化法によって推定し、補正した第１の遺伝子発現強度補正データを送出するスポット位置による第１の補正手段と、前記第１の遺伝子発現強度補正データに対してＳ－Ｄ変換を行い、遺伝子発現強度データに潜在しうる蛍光色素の感度の違いによる歪みをノンパラメトリック平滑化法によって推定し、蛍光色素の感度の違いによる歪みが補正された第２の遺伝子発現強度補正データを送出する第２の補正手段とを具備するデータ解析装置と、

前記第２の遺伝子発現強度補正データを出力する出力装置を

有することを特徴とするｃＤＮＡマイクロアレイデータの補正システム。

【請求項２】　さらに、任意の段階で遺伝子発現強度データの歪みを定量化し、Ｓ－Ｄプロット上に視覚化するＳ－Ｄ変換手段を有していることを特徴とする請求項１記載のｃＤＮＡマイクロアレイデータの補正システム。

【請求項３】　前記順序統計量は以下の数１（尚、外１は、前記規準化遺伝子発現強度データであり、外２はチャンネルによって得られた全スポットの遺伝子発現強度データであり、$L_k(c)$および$M_k(c)$はそれぞれグリッドｋにおいて、チャンネルｃによって得られた遺伝子発現強度データの２５％点および５０％点を示す。）で示されることを特徴とする請求項１又は２記載のｃＤＮＡマイ

クロアレイデータの補正システム。

【数１】

$$w_{ij}^{k}(c)=\frac{y_{ij}^{k}(c)-L_{k}(c)}{M_{k}(c)-L_{k}(c)},\quad c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J,\ k=1,\cdots,K.$$

【外１】

$w_{ij}^{k}(c)$

【外２】

$y_{ij}^{k}(c)$

【請求項４】　前記データ規準化手段は、少なくとも２つの遺伝子発現強度データチャンネルによって得られた全スポットの遺伝子発現強度データを規準化したかどうかを判定し、全スポットの遺伝子発現強度データを規準化するまで続けることを特徴とする請求項３記載のｃＤＮＡマイクロアレイデータの補正システム。

【請求項５】　前記規準化遺伝子発現強度データは、真の発現強度とスポット位置に依存する歪みとの和によって表されることを特徴とする請求項１記載のｃＤＮＡマイクロアレイデータの補正システム。

【請求項６】　前記第１の補正手段は、スポット位置に依存する歪みをｘ軸、ｙ軸、及び前記ｘ，ｙ軸の交互作用による歪み（それぞれ、外３，外４，外５とする。）の回帰関係で示されるノンパラメトリック回帰モデルにより記述し、以下の数２に示されるノンパラメトリック平滑化法を用いて、スポット位置による歪み（外６）を推定することを特徴とする請求項１記載のｃＤＮＡマイクロアレイデータの補正システム。

【数２】

$$\hat{\xi}^{k}_{ij}(c)=\hat{\alpha}^{(c)}_{k}(i)+\hat{\beta}^{(c)}_{k}(j)+\hat{\gamma}^{(c)}_{k}((i-m_i)(j-m_j)),\ c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J.$$

【外３】

$\alpha^{(c)}_{k}(i)$

【外４】

$\beta^{(c)}_{k}(j)$

【外５】

$\gamma^{(c)}_{k}((i-m_i)(j-m_j))$

【外６】

$\xi^{k}_{ij}(c)$

【請求項７】　前記スポット位置による歪みの補正は、以下の数３（尚、外７は補正された真の遺伝子発現強度データである。）に従って行なわれることを特徴とする請求項６記載のｃＤＮＡマイクロアレイデータの補正システム。

【数３】

$$\hat{z}^{k}_{ij}(c) = w^{k}_{ij}(c) - \hat{\xi}^{k}_{ij}(c)$$

【外７】

$\hat{z}^{k}_{ij}(c)$

【請求項８】　前記第２の補正手段における前記Ｓ－Ｄ変換は、以下の数４に従って行なわれることを特徴とする請求項７記載のｃＤＮＡマイクロアレイデータの補正システム。

【数４】

$$u_{ij}^{k} = \hat{z}_{ij}^{k}(1) + \hat{z}_{ij}^{k}(2)$$

$$v_{ij}^{k} = \hat{z}_{ij}^{k}(1) - \hat{z}_{ij}^{k}(2)$$

【請求項９】　前記第２の補正手段は、以下の数５で示されるノンパラメトリック回帰モデルにより記述し、以下の数６及び数７で示されるノンパラメトリック平滑化法を用いて、蛍光色素の感度による測定誤差を推定し、補正を行うことを特徴とする請求項８記載のｃＤＮＡマイクロアレイデータの補正システム。

【数５】

$$v_{ij}^{k} = \phi\left(u_{ij}^{k}\right) + \varepsilon_{ij}^{k} \quad , \varepsilon_{ij}^{k} = N\left(0, v^{2}\right)$$

【数６】

$$\eta_{ij}^{k} = v_{ij}^{k} - \hat{\phi}\left(u_{ij}^{k}\right)$$

【数７】

$$\hat{y}_{ij}^{k}(1) = \frac{1}{2}\left(u_{ij}^{k} + \eta_{ij}^{k}\right)$$

$$\hat{y}_{ij}^{k}(2) = \frac{1}{2}\left(u_{ij}^{k} - \eta_{ij}^{k}\right)$$

【請求項１０】　前記補正の前提として、遺伝子が発現している確率は０．５より小さいと仮定し、各グリッド内の半分以上のスポットで検出される蛍光強度は、バックグラウンドノイズあるいは系統誤差を示しているとすることを特徴とする請求項１記載のｃＤＮＡマイクロアレイデータの補正システム。

【請求項１１】　さらに前記補正の前提として、グリッドにおいて、少なくとも２つの遺伝子発現強度データチャンネルによって得られた蛍光強度の２５％点と５０％点を、それぞれ$L_k(c)$および$M_k(c)$とするとき、遺伝子の大半は非発現状態にあり全てのグリッドとチャンネルにおいて蛍光強度の５０％点以下の分布は共通であるという前提に基づき、$L_k(c)$と$M_k(c)-L_k(c)$は各グリッドおよび各チャンネルで等しいと仮定することを特徴とする請求項１０記載のｃＤＮＡマイクロアレイデータの補正システム。

【請求項１２】　マイクロアレイデータのグローバル及びローカルな歪みに

対してより精密な補正をし、さらに蛍光色素の感度の違いによる測定誤差を補正するｃＤＮＡマイクロアレイデータの補正方法において、

各スポットのバックグラウンドノイズの除去及び信頼性を示すフラッグ情報を考慮し、あらかじめ調整されている遺伝子発現強度データを入力するステップと、　大半の遺伝子は発現していないことを前提として、前記入力された遺伝子発現強度データに対してグリッド毎の順序統計量を用いて、当該遺伝子発現強度データを規準化するステップと、

前記規準化された規準化遺伝子発現強度データを出力するステップと、

前記規準化遺伝子発現強度データに対して、グリッドの座標におけるスポット位置に依存する歪みをノンパラメトリック平滑化法によって推定し、スポット位置に依存したデータの歪みを補正するステップと、

前記スポット位置に依存したデータ歪みの補正がされた第１の遺伝子発現強度補正データを出力するステップと、

前記第１の遺伝子発現強度補正データに対して、Ｓ－Ｄ変換を行い、遺伝子発現強度データに潜在しうる蛍光色素の感度の違いによる歪みをノンパラメトリック平滑化法によって推定し、蛍光色素の感度の違いによる歪みを補正するステップと、

前記蛍光色素の感度の違いによる歪みの補正がされた第２の遺伝子発現強度補正データを出力するステップとを

有することを特徴とするｃＤＮＡマイクロアレイデータの補正方法。

【請求項１３】　さらに、任意の段階で遺伝子発現強度データの歪みを定量化し、Ｓ－Ｄプロット上に視覚化するステップを有していることを特徴とする請求項１２記載のｃＤＮＡマイクロアレイデータの補正方法。

【請求項１４】　前記順序統計量は以下の数８（尚、外８は、前記規準化遺伝子発現強度データであり、外９はチャンネルによって得られた全スポットの遺伝子発現強度データであり、$L_k(c)$および$M_k(c)$はそれぞれグリッドｋにおいて、チャンネルｃによって得られた遺伝子発現強度データの２５％点および５０％点を示す。）で示されることを特徴とする請求項１２又は１３記載のｃＤＮＡマイクロアレイデータの補正方法。

【数８】

$$w_{ij}^{k}(c)=\frac{y_{ij}^{k}(c)-L_{k}(c)}{M_{k}(c)-L_{k}(c)},\quad c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J,\ k=1,\cdots,K.$$

【外８】

$w_{ij}^{k}(c)$

【外９】

$y_{ij}^{k}(c)$

【請求項１５】　前記データを規準化するステップにおいて、少なくとも２つの遺伝子発現強度データチャンネルによって得られた全スポットの遺伝子発現強度データを規準化したかどうかを判定し、全スポットの遺伝子発現強度データを規準化するまで続けることを特徴とする請求項１４記載のｃＤＮＡマイクロアレイデータの補正方法。

【請求項１６】　前記規準化遺伝子発現強度データは、真の発現強度とスポット位置に依存する歪みとの和によって表されることを特徴とする請求項１５記載のｃＤＮＡマイクロアレイデータの補正方法。

【請求項１７】　前記スポット位置に依存したデータの歪みを補正するステップにおいて、スポット位置に依存する歪みをｘ軸、ｙ軸、及び前記ｘ，ｙ軸の交互作用による歪み（それぞれ、外１０，外１１，外１２とする。）の回帰関係で示されるノンパラメトリック回帰モデルにより記述し、以下の数９で示されるノンパラメトリック平滑化法を用いて、スポット位置による歪み（外１３）を推定することを特徴とする請求項１２記載のｃＤＮＡマイクロアレイデータの補正方法。

【数９】

$$\hat{\xi}^{k}_{ij}(c)=\hat{\alpha}^{(c)}_{k}(i)+\hat{\beta}^{(c)}_{k}(j)+\hat{\gamma}^{(c)}_{k}((i-m_i)(j-m_j))\ ,\ c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J\ .$$

【外１０】

$\alpha^{(c)}_{k}(i)$

【外１１】

$\beta^{(c)}_{k}(j)$

【外１２】

$\gamma^{(c)}_{k}((i-m_i)(j-m_j))$

【外１３】

$\xi^{k}_{ij}(c)$

【請求項１８】　前記スポット位置による歪みの補正は、以下の数１０（尚、外１４は補正された真の遺伝子発現強度データである。）に従って行なわれることを特徴とする請求項１７記載のｃＤＮＡマイクロアレイデータの補正方法。

【数１０】

$$\hat{z}^{k}_{ij}(c)\ =\ w^{k}_{ij}(c)\ -\ \hat{\xi}^{k}_{ij}(c)$$

【外１４】

$\hat{z}^{k}_{ij}(c)$

【請求項１９】　前記蛍光色素の感度の違いによる歪みを補正するステップにおける前記Ｓ－Ｄ変換は、以下の数１１に従って行なわれることを特徴とする請求項１８記載のｃＤＮＡマイクロアレイデータの補正方法。

【数１１】

$$u_{ij}^{k} = \hat{z}_{ij}^{k}(1) + \hat{z}_{ij}^{k}(2)$$

$$v_{ij}^{k} = \hat{z}_{ij}^{k}(1) - \hat{z}_{ij}^{k}(2)$$

【請求項２０】　前記蛍光色素の感度の違いによる歪みを補正するステップにおいて、以下の数１２で示されるノンパラメトリック回帰モデルにより記述し、以下の数１３及び数１４で示されるノンパラメトリック平滑化法を用いて、蛍光色素の感度による測定誤差を推定し、補正を行うことを特徴とする請求項１９記載のｃＤＮＡマイクロアレイデータの補正方法。

【数１２】

$$v_{ij}^{k} = \phi\left(u_{ij}^{k}\right) + \varepsilon_{ij}^{k} \quad , \varepsilon_{ij}^{k} = N\left(0, v^{2}\right)$$

【数１３】

$$\eta_{ij}^{k} = v_{ij}^{k} - \hat{\phi}\left(u_{ij}^{k}\right)$$

【数１４】

$$\hat{y}_{ij}^{k}(1) = \frac{1}{2}\left(u_{ij}^{k} + \eta_{ij}^{k}\right)$$

$$\hat{y}_{ij}^{k}(2) = \frac{1}{2}\left(u_{ij}^{k} - \eta_{ij}^{k}\right)$$

【請求項２１】　前記補正の前提として、遺伝子が発現している確率は０．５より小さいと仮定し、各グリッド内の半分以上のスポットで検出される蛍光強度は、バックグラウンドノイズあるいは系統誤差を示しているとすることを特徴とする請求項１２記載のｃＤＮＡマイクロアレイデータの補正方法。

【請求項２２】　さらに前記補正の前提として、グリッドにおいて、少なくとも２つの遺伝子発現強度データチャンネルによって得られた蛍光強度の２５％点と５０％点を、それぞれ$L_k(c)$および$M_k(c)$とするとき、遺伝子の大半は非発現状態にあり、全てのグリッドとチャンネルにおいて蛍光強度の５０％点以下の分布は共通であるという前提に基づき、$L_k(c)$と$M_k(c) - L_k(c)$は各グリッドおよび各チャンネルで等しいと仮定することを特徴とする請求項２１記載のｃＤＮＡマイクロアレイデータの補正方法。

【請求項２３】　マイクロアレイデータのグローバル及びローカルな歪みに対してより精密な補正をし、さらに蛍光色素の感度の違いによる測定誤差を補正するためコンピュータに、

各スポットのバックグラウンドノイズの除去及び信頼性を示すフラッグ情報を考慮し、あらかじめ調整されている遺伝子発現強度データを入力するステップと、　大半の遺伝子は発現していないことを前提として、前記入力された遺伝子発現強度データに対してグリッド毎の順序統計量を用いて、当該遺伝子発現強度データを規準化するステップと、

前記規準化された規準化遺伝子発現強度データを出力するステップと、

前記規準化遺伝子発現強度データに対して、グリッドの座標におけるスポット位置に依存する歪みをノンパラメトリック平滑化法によって推定し、スポット位置に依存したデータの歪みを補正するステップと、

前記スポット位置に依存したデータ歪みの補正がされた第１の遺伝子発現強度補正データを出力するステップと、

前記第１の遺伝子発現強度補正データに対して、Ｓ－Ｄ変換を行い、遺伝子発現強度データに潜在しうる蛍光色素の感度の違いによる歪みをノンパラメトリック平滑化法によって推定し、蛍光色素の感度の違いによる歪みを補正するステップと、

前記蛍光色素の感度の違いによる歪みの補正がされた第２の遺伝子発現強度補正データを出力するステップとを

を実行させるためのｃＤＮＡマイクロアレイデータ補正プログラム。

【請求項２４】　マイクロアレイデータのグローバル及びローカルな歪みに対してより精密な補正をし、さらに蛍光色素の感度の違いによる測定誤差を補正するためコンピュータに、

各スポットのバックグラウンドノイズの除去及び信頼性を示すフラッグ情報を考慮し、あらかじめ調整されている遺伝子発現強度データを入力するステップと、　大半の遺伝子は発現していないことを前提として、前記入力された遺伝子発現強度データに対してグリッド毎の順序統計量を用いて、当該遺伝子発現強度データを規準化するステップと、

前記規準化された規準化遺伝子発現強度データを出力するステップと、

前記規準化遺伝子発現強度データに対して、グリッドの座標におけるスポット位置に依存する歪みをノンパラメトリック平滑化法によって推定し、スポット位置に依存したデータの歪みを補正するステップと、

前記スポット位置に依存したデータ歪みの補正がされた第１の遺伝子発現強度補正データを出力するステップと、

前記第１の遺伝子発現強度補正データに対して、Ｓ－Ｄ変換を行い、遺伝子発現強度データに潜在しうる蛍光色素の感度の違いによる歪みをノンパラメトリック平滑化法によって推定し、蛍光色素の感度の違いによる歪みを補正するステップと、

前記蛍光色素の感度の違いによる歪みの補正がされた第２の遺伝子発現強度補正データを出力するステップとを

を実行させるためのｃＤＮＡマイクロアレイデータ補正プログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、数理モデルに基づいたｃＤＮＡマイクロアレイデータのデータ補正システム、方法、プログラム及び記録媒体に関し、特にグローバルノーマライゼーションとローカルノーマライゼーション、さらに蛍光色素の感度の違いによる測定の歪みの補正をすることができるｃＤＮＡマイクロアレイデータの補正システム、方法、プログラム及び記録媒体に関するものである。

【０００２】

【従来の技術】

現在、ゲノム研究は個々の遺伝子についての構造解析から体系的な遺伝子の機能解析へと展開しつつある。機能未知の遺伝子や総体としての遺伝子の機能解析のために、多数の遺伝子の発現強度を同時に定量化することのできるｃＤＮＡ（相補的なＤＮＡ）マイクロアレイを用いた実験はその有効性が大いに期待されている。

【０００３】

二色蛍光法によるｃＤＮＡマイクロアレイを用いた実験の目的は二種類の細胞の遺伝子発現の違いを検出することにある。ここで、二色蛍光法によるｃＤＮＡマイクロアレイの概要について述べる。まず、多数の遺伝子セットのｃＤＮＡを参照用のプローブとして、スライドグラス上にアレイ状に高密度に固定化する（マイクロアレイ）。

【０００４】

次に、条件の異なる２種類のサンプル、細胞１と細胞２（例えば正常細胞と癌細胞）から抽出したｍＲＮＡをそれぞれ波長の異なる蛍光色素でラベルし、ターゲットｃＤＮＡを合成する。そして、それらを等量混合したものをマイクロアレイに固定化された参照用のプローブｃＤＮＡに競合的にハイブリダイズさせる。ハイブリダイゼーション後、スキャナーでそれぞれの蛍光色素強度を測定する。細胞１にラベルされた蛍光色素をチャンネル1により、細胞２にラベルされた蛍光色素をチャンネル２により読み取り、それぞれを各細胞の遺伝子発現強度データ（マイクロアレイデータ）とする。

【０００５】

このように、マイクロアレイデータが得られるまでの過程は複雑であり、高度な実験技術が必要とされることから、実験の各段階において様々な実験誤差が生じると考えられる。このため、マイクロアレイデータから真に生物学的意味のあるデータを取り出すためには遺伝子発現強度の分布と実験誤差の解析は解決すべき重要な課題である。

【０００６】

遺伝子発現強度の分布に関しては、例えば、以下の非特許文献１を参照すると、Ｎｅｗｔｏｎ等は遺伝子発現強度にガンマ分布関数を仮定し、遺伝子発現強度比（チャンネル１とチャンネル２の遺伝子発現強度データの比）についての統計学的性質を考察している。

【０００７】

また、観測された遺伝子発現強度データに対しては、例えば、以下の非特許文献２を参照すると、Ｌｅｅ等は真の遺伝子発現強度を２個の水準値に分離できる

ことおよび偶然誤差の存在を前提として、以下の数１５に示されるような混合正規分布を適用し、遺伝子発現強度データについての統計学的考察を行った。

【０００８】

【数１５】

$$f(x) = p\phi\left(x - \mu_1 | \sigma_1^2\right) + (1-p)\phi\left(x - \mu_2 | \sigma_2^2\right)$$

ここで、ｘはスキャナーなどによって得られる蛍光強度などの遺伝子発現強度データを表し、右辺第１項の外１５は

【外１５】

$$\phi\left(x - \mu_1 | \sigma_1^2\right)$$

遺伝子が発現しているときの平均μ_1、分散外１６の正規分布、

【外１６】

$$\sigma_1^2$$

また、同第２項の外１７は遺伝子が発現していないときの平均μ_2、分散外１８の正規分布の密度関数を表し、

【外１７】

$$\phi\left(x - \mu_2 | \sigma_2^2\right)$$

【外１８】

$$\sigma_2^2$$

ｐはその混合率を表す母数である。

【０００９】

実験誤差の解析については、系統誤差の除去、いわゆるノーマライゼーション

の方法がいくつか提案されている。ノーマライゼーションの方法は、大きく分けてアレイ上のすべてのスポットを対象にしたグローバルノーマライゼーションと，あるサブセットに分けた（例えばグリッド単位の）スポットを対象にしたローカルノーマライゼーションの二つが提案されている。グローバルノーマライゼーションについては、例えば、以下の非特許文献３を参照すると、Ｃｈｅｎ等は二つの細胞の遺伝子発現強度の中央値は等しいとしてチャンネル１とチャンネル２で得られた測定値の補正を行った。ローカルノーマライゼーションについては、例えば、以下の非特許文献４、５、６を参照すると、ＤｕｄｏｉｔやＳｃｈｕｃｈｈａｒｄｔやＹａｎｇは、系統誤差が、スポットのスライドグラス上の位置や、二種類の蛍光色素の感度の違いによって生じたものと考え、それらを除去する方法を提案した。

【００１０】

【非特許文献１】

Newton et. al、2001年、ジャーナル・オブ・コンピュテーショナル・バイオロジー、第8巻、37～52頁（Journal of Computational Biology Vol. 8, pp. 37-52）

【００１１】

【非特許文献２】

Lee et. al、2000年、プロシーディング・オブ・ザ・ナショナル・アカデミィー・オブ・サイエンシズ、第97巻、第18号、9834～9839頁（Proceeding of the National Academy of Sciences Vol. 97, No 18, pp. 9834-9839）

【００１２】

【非特許文献３】

Chen et. al、1997年、ジャーナル・オブ・バイオメディカル・オプティクス、第２号、364~374頁（Journal of Biomedical Optics Vol. 2, pp. 364-374）

【００１３】

【非特許文献４】

Dudoit et. al、2000. Statistical methods for identifying differentia

lly expressed genes in replicated cDNA microarray experiments. Technical ~Report #578 2.

【００１４】

【非特許文献５】

Schuchhardt et. al、2000年、ヌクレ・アシッド・リサーチ、第28巻、第10号（Nucleic Acids Research, 2000, Vol.28, No. 10）

【００１５】

【非特許文献６】

Yang et. al、 2002年、ヌクレ・アシッド・リサーチ、第30巻、第4号（Nucleic Acids Research, 2002, Vol.30, No. 4）

【００１６】

【発明が解決しようとする課題】

上記した従来技術における問題点は、マイクロアレイデータの解析結果は再現性に乏しく不安定なものになりがちで、精度や効率は低いものとみなされていることである。その理由は、遺伝子の発現に関する真の信号と実験誤差の分離が十分に行われていないからである。その背景要因として、遺伝子の発現強度はそれぞれの遺伝子によってレベルが異なっていることが考えられ、その場合、上記数１５によるモデルは明らかに過大に単純化されすぎている。

【００１７】

本発明の目的は、マイクロアレイ上の遺伝子発現強度データに関してよりもっともな数理モデルを想定して、グローバルおよびローカルな歪みに対して高い精度の補正を行い、さらに蛍光色素の感度の違いによる測定誤差を補正するための包括的なノーマライゼーションの方法およびシステムを提供することである。

【００１８】

【課題を解決するための手段】

本発明のｃＤＮＡマイクロアレイデータの補正システムは、蛍光強度などの遺伝子発現強度データを入力する入力装置と、プログラムの制御により動作するデータ解析装置と、出力装置とを含む。なお，入力される遺伝子発現強度データは、各スポットのバックグラウンドノイズの除去や各スポットの信頼性を示すフラ

ッグ情報を考慮し、あらかじめ調整されているものとする。

【００１９】

前記データ解析装置は、下記の三個の連続した処理過程で構成される。第一処理過程であるデータ規準化手段では、前記入力装置から遺伝子発現強度データを入力し、大半の遺伝子は発現していないことを前提としてグリッド毎の順序統計量を用いて遺伝子発現強度データを規準化し、規準化した遺伝子発現強度データを出力する。

【００２０】

第二処理過程であるスポット位置による補正手段では、前記規準化された遺伝子発現強度データを入力し、グリッド毎にスポット位置に依存する歪みをノンパラメトリック平滑化法によって推定し、スポット位置に依存したデータの歪みを補正した遺伝子発現強度データを出力する。

【００２１】

第三処理過程であるＳ－Ｄプロットによる補正手段では、第二処理過程の段階まで補正された遺伝子発現強度データに対してMA変換の変形であるＳ－Ｄ変換(ＭＡ変換およびＭＡプロットについては、上記非特許文献６を参照)を行い、遺伝子発現強度データに潜在しうる蛍光色素の感度の違いによる歪みをノンパラメトリック平滑化法によって推定し、蛍光色素の感度の違いによる歪みを補正した遺伝子発現強度データを前記出力装置に出力する。

【００２２】

なお、本システムは、任意の段階で遺伝子発現強度データの歪みを定量化し、Ｓ－Ｄプロット上に視覚化するＳ－Ｄ変換手段を有していることを特徴とする。

【００２３】

このような構成を採用し、遺伝子発現強度データを補正することにより、本発明の目的を達成することができる。

【００２４】

【発明の実施の形態】

はじめに、本発明におけるマイクロアレイの構造を説明する。図１を参照すると、Ｋ個の各グリッドにＩ×Ｊ個ずつ、合計Ｋ×Ｉ×Ｊ個のｃＤＮＡがスライド

グラス上にスポットされている。いま、グリッドｋにおける座標（ｉ，ｊ）にスポットされたｃＤＮＡに対して、チャンネルｃ＝１，２によって得られた蛍光強度を外１９とする。

【００２５】

【外１９】

$$y_{ij}^{k}(c),\ c = 1,2$$

次に、以下の２つの仮定をする。

【００２６】

（仮定１）

遺伝子が発現している確率は0.5より小さいと仮定し、各グリッド内の半分以上のスポットで検出される蛍光強度外２０は、バックグラウンドノイズあるいは系統誤差を示しているとする。

【００２７】

【外２０】

$$y_{ij}^{k}(c)$$

（仮定２）

グリッドｋにおいて、チャンネルｃによって得られた蛍光強度外２１の２５％点と５０％点を、

【外２１】

$$y_{ij}^{k}(c)$$

それぞれ$L_k(c)$および$M_k(c)$とするとき、遺伝子の大半は非発現状態にあり全てのグリッドとチャンネルにおいて蛍光強度の５０％点以下の分布は共通であるという前提に基づき、$L_k(c)$と$M_k(c)-L_k(c)$は各グリッドおよび各チャンネルで等しいと仮定する。

【００２８】

次に、以上の仮定をもとに、本発明の第１の実施の形態について図面を参照して詳細に説明する。図２を参照すると、本発明の第１の実施の形態は、蛍光強度などの遺伝子発現強度データを入力する入力装置１と、プログラム制御により動作するデータ解析装置２と、ディスプレイ装置や印刷装置等の出力装置３とを含む。データ解析装置は、データ規準化手段２１と、スポット位置による補正手段２２と、Ｓ－Ｄプロットによる補正手段２３とを備えている。

【００２９】

データ規準化手段２１は、与えられた遺伝子発現強度データに対して、グリッド毎の順序統計量を用いて遺伝子発現強度データを規準化し、スポット位置による補正手段２２及びＳ－Ｄ変換手段２４に送る。

【００３０】

スポット位置による補正手段２２は、データ規準化手段２１から送られてきた規準化された遺伝子発現強度データに対して、グリッド毎にスポット位置に依存する歪みをノンパラメトリック平滑化法によって推定し、補正した遺伝子発現強度データをＳ－Ｄプロットによる補正手段２３及びＳ－Ｄ変換手段２４に送る。

【００３１】

Ｓ－Ｄプロットによる補正手段２３は、スポット位置による補正手段２２から送られてきた補正された遺伝子発現強度データにＳ－Ｄ変換を行い、蛍光色素の感度の違いに起因する歪みをノンパラメトリック平滑化法により補正した後、遺伝子発現強度データを出力装置３へ送る。

【００３２】

Ｓ－Ｄ変換手段２４は送られてきた遺伝子発現強度データにＳ－Ｄ変換を行い、出力装置３へ送る。

【００３３】

次に、図２、図３を参照して本実施の形態について詳細に説明する。入力装置１より入力された蛍光強度などの遺伝子発現強度データはデータ規準化手段２１へ送られる。データ規準化手段２１は、送られてきた発現強度データに対して、以下の数１６で示されるように、グリッド毎の順序統計量を用いて発現強度デー

タを規準化する（図３のステップＡ１）。

【００３４】

【数１６】

$$w_{ij}^{k}(c)=\frac{y_{ij}^{k}(c)-L_{k}(c)}{M_{k}(c)-L_{k}(c)},\quad c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J,\ k=1,\cdots,K.$$

２つのチャンネルによって得られた全スポットの遺伝子発現強度データ外２２を規準化したかどうかを判定し、

【外２２】

$y_{ij}^{k}(c)$

全スポットの遺伝子発現強度データ（２×Ｉ×Ｊ×Ｋ個）を規準化するまで続ける（ステップＡ２）。

【００３５】

データ規準化手段２１において規準化された遺伝子発現強度データ外２３に対して、

【外２３】

$w_{ij}^{k}(c)$

外２４を真の発現強度を反映した蛍光強度（以下、真の発現蛍光強度）とし、

【外２４】

$z_{ij}^{k}(c)$

外２５をグリッドｋの座標（ｉ，ｊ）におけるスポット位置に依存する歪みとする。

【００３６】

【外２５】

$\xi_{ij}^{k}(c)$

このとき、以下の数１７に示すように、遺伝子発現強度データ外２６は、真の発現強度外２７とスポット位置に依存する歪み外２８との和によって表されるとする。

【００３７】

【外２６】

$w_{ij}^{k}(c)$

【外２７】

$z_{ij}^{k}(c)$

【外２８】

$\xi_{ij}^{k}(c)$

【数１７】

$$w_{ij}^{k}(c) = z_{ij}^{k}(c) + \xi_{ij}^{k}(c) + \varepsilon_{ij}^{k}(c),\ \varepsilon_{ij}^{k}(c) \sim N\left(0,\ \sigma_{k}(c)^{2}\right),\ c = 1,2.$$

ただし、外２９はランダムなノイズであるとする。

【００３８】

【外２９】

$\varepsilon_{ij}^{k}(c)$

スポット位置による補正手段２２は、以下の数１８に示すようにスポット位置

に依存する歪み外３０を「ｘ軸」、「ｙ軸」および「２つの軸の交互作用」による歪みの回帰関係で示されるノンパラメトリック回帰モデルにより記述し、

【外３０】

$$\xi_{ij}^{k}(c)$$

以下の数１９に示すようにノンパラメトリック平滑化法を用いて、スポット位置による歪み外３１を推定する。

【００３９】

【外３１】

$$\xi_{ij}^{k}(c)$$

【数１８】

$$\xi_{ij}^{k}(c)=\alpha_{k}^{(C)}(i)+\beta_{k}^{(C)}(j)+\gamma_{k}^{(C)}((i-m_i)(j-m_j)),\ c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J,$$

$$\sum_i \alpha_{k}^{(C)}(i)=0,\ \sum_j \beta_{k}^{(C)}(j)=0,\ \sum_i \sum_j \gamma_{k}^{(C)}((i-m_i)(j-m_j))=0.$$

【数１９】

$$\hat{\xi}_{ij}^{k}(c)=\hat{\alpha}_{k}^{(C)}(i)+\hat{\beta}_{k}^{(C)}(j)+\hat{\gamma}_{k}^{(C)}((i-m_i)(j-m_j)),\ c=1,2,\ i=1,\cdots,I,\ j=1,\cdots,J.$$

ここで、外３２とする。

【００４０】

【外３２】

$$m_i=\lfloor I/2\rfloor,\ m_j=\lfloor J/2\rfloor$$

外３３はα以上の最小の整数とする。

【００４１】

【外３３】

⌊α⌋

スポット位置による補正手段２２は、以下の数２０に示すように、データ規準化手段２１において規準化された遺伝子発現強度データ外３４に対して、推定されたスポット位置による歪み外３５を補正する（ステップＡ３）。

【００４２】

【外３４】

$w_{ij}^{k}(c)$

【外３５】

$\hat{\xi}_{ij}^{k}(c)$

【数２０】

$$\hat{z}_{ij}^{k}(c) = w_{ij}^{k}(c) - \hat{\xi}_{ij}^{k}(c)$$

データ規準化手段２１において規準化された全スポットの遺伝子発現強度データ外３６に対して、

【外３６】

$w_{ij}^{k}(c)$

スポット位置による歪み外３７の補正をしたかどうかを判定し、

【外３７】

$\hat{\xi}_{ij}^{k}(c)$

全スポットの遺伝子発現強度データ（２×Ｉ×Ｊ×Ｋ個）を補正するまで続ける（ステップＡ４）。

【００４３】

Ｓ－Ｄプロットによる補正手段２３は、スポット位置による補正手段２２において補正された真の遺伝子発現強度データ外３８に対して、

【外３８】

$\hat{z}^{k}_{ij}(c)$

以下の数２１に示すように、Ｓ－Ｄ変換を行う。

【００４４】

【数２１】

$$u^{k}_{ij} = \hat{z}^{k}_{ij}(1) + \hat{z}^{k}_{ij}(2)$$
$$v^{k}_{ij} = \hat{z}^{k}_{ij}(1) - \hat{z}^{k}_{ij}(2)$$

さらに、以下の数２２で示されるようなノンパラメトリック回帰モデルを記述し、以下の数２３及び数２４に示すようにノンパラメトリック平滑化法を用いて蛍光色素の感度による測定誤差を推定し、補正を行う（ステップＡ５）。

【００４５】

【数２２】

$$v^{k}_{ij} = \phi\left(u^{k}_{ij}\right) + \varepsilon^{k}_{ij} \quad , \varepsilon^{k}_{ij} = N\left(0, v^{2}\right)$$

【数２３】

$$\eta^{k}_{ij} = v^{k}_{ij} - \hat{\phi}\left(u^{k}_{ij}\right)$$

【数２４】

$$\hat{y}^{k}_{ij}(1) = \frac{1}{2}\left(u^{k}_{ij} + \eta^{k}_{ij}\right)$$
$$\hat{y}^{k}_{ij}(2) = \frac{1}{2}\left(u^{k}_{ij} - \eta^{k}_{ij}\right)$$

スポット位置による補正手段２２において補正された真の遺伝子発現強度データ外３９に対して、

【外３９】

$\hat{z}^{k}_{ij}(c)$

Ｓ－Ｄプロットによる補正をしたかどうかを判定し、全スポットの真の遺伝子発現強度データ（２×Ｉ×Ｊ×Ｋ個）を補正するまで続ける（ステップＡ６）。

【００４６】

なお、図３のＡ２、Ａ４の各ステップ終了後、遺伝子発現強度データはＳ－Ｄ変換手段２４を介して出力装置３に送られ、Ｓ－Ｄプロットによって遺伝子発現強度データの歪みを視覚化することができる。

【００４７】

次に、本実施の形態の効果について説明する。本実施の形態では、グリッド間での順序統計量を用いた規準化（グローバルノーマライゼーション）とグリッド内でのスポット位置に依存する歪みの補正（ローカルノーマライゼーション）を組み合わせたノーマライゼーションを行った。これにより、グリッド間での遺伝子発現強度の偏りによる系統誤差と、グリッド内でのスポット位置に依存する歪みを同時に補正することができる。さらに、Ｓ－Ｄプロットによる補正においては、発現強度データの和と差を用いることにより、蛍光色素の感度の違いによる測定誤差を補正することができる。

【００４８】

次に、本発明の第２の実施の形態について図面を参照して詳細に説明する。図４を参照すると、本発明の第２の実施の形態は、本発明の第１の実施の形態と同様に、入力装置、データ解析装置、出力装置を備え、更に、データ解析プログラムを記録した記録媒体４を備える。この記録媒体４は可搬形あるいは固定型のいずれであってもよく、磁気ディスク、半導体メモリ、ＣＤ－ＲＯＭその他の記録媒体であってもよい。

【００４９】

また、本手法を実行できるコンピュータプログラムを、ネットワークに接続されたコンピュータの記録装置に格納しておき、ネットワークを介して他のコンピュータに転送することもできる。本アルゴリズムを実行するコンピュータプログラムを提供する提供媒体としては、様々な形式のコンピュータに読み出し可能な媒体として頒布可能であって、特定のタイプの媒体に限定されるものではない。データ解析プログラムは記録媒体４からデータ解析装置５に読み込まれ、データ

解析装置２の動作を制御し、入力装置１から入力されたデータファイルに対して第1の実施の形態におけるデータ処理装置２による処理と同一の処理を実行する。

【００５０】

【実施例】

以下、本発明の実施例について説明する。例として用いたデータは，異なる２種類の癌細胞（Ａ細胞、Ｂ細胞）の遺伝子発現状況の比較のために行われた実験から得られたものである。

【００５１】

一枚のチップ上に４８グリッド，１グリッドあたり４４１（２１×２１）スポット，計２１１６８の遺伝子の発現パターンについて調べたものである。

【００５２】

図５、図７はチャンネル１により得られたオリジナルデータのＡ細胞遺伝子発現強度を示し、図６、図８はチャンネル２によって得られたオリジナルデータのＢ細胞遺伝子発現強度を示す。それぞれの図は、マイクロアレイ上のスポット位置に対する遺伝子発現強度の対数値をプロットしたものである。また、図７、図８は第１グリッドから第４グリッドまでを拡大したものである。図５～図８を見ると、遺伝子発現強度がグリッドごとに周期的に繰り返される系統的な歪みが観察される。マイクロアレイ上の遺伝子は無作為にスポットされているので、このような歪みは実験誤差と考えられる。

【００５３】

図９は，それらのＳ－Ｄプロットである。横軸は，各チャンネルの遺伝子発現強度の和，縦軸はそれらの差をとったものを示している。各チャンネルの遺伝子発現強度の和が小さい領域と大きい領域においては、各チャンネルの遺伝子発現強度の差は真の遺伝子発現の違いによる影響は小さく、各チャンネルの蛍光色素の感度の違いによるものと考えられる。これにより、図９において蛍光色素の感度の違いによって生じたと考えられる歪みが観察される。

【００５４】

図１０に、チャンネル１におけるオリジナルデータのスポット位置に対する遺

伝子発現強度の図を示す。図１１に、チャンネル１における第一処理過程後のスポット位置に対する遺伝子発現強度の図を示す。図１２に、チャンネル１における第二処理過程後のスポット位置に対する遺伝子発現強度の図を示す。スポット位置に依存していたグリッドごとに周期的に繰り返される系統的な歪みが補正されて取り除かれていることがわかる。

【００５５】

図１３にチャンネル１における第三処理過程後のスポット位置に対する遺伝子発現強度の図を示す。図１４～図１７にチャンネル２におけるオリジナルデータ、第一処理過程後、第二処理過程後、第三処理過程後のスポット位置に対する遺伝子発現強度の図を示す。チャンネル１と同様にスポット位置に依存していたグリッドごとに周期的に繰り返される系統的な歪みが補正されて取り除かれていることがわかる。

【００５６】

図１８～図２１にオリジナルデータ、第一処理過程後、第二処理過程後、第三処理過程後のＳ－Ｄプロットを示す。図２１を見ると、蛍光色素の感度の違いによる歪みが補正されて取り除かれていることがわかる。

【００５７】

【発明の効果】

本発明によれば、グリッド間での位置および尺度の揺らぎに対する頑健な順序統計量の２５％点と５０％点による規準化（グローバルノーマライゼーション）と、グリッド内でのスポット位置に依存する歪みの補正（ローカルノーマライゼーション）を組み合わせてノーマライゼーションを行っているため、グリッド間での遺伝子発現強度の偏りや感度の揺らぎによる系統誤差と、グリッド内でのスポット位置に依存する歪みを、発現している遺伝子の頻度や外れ値の影響をほとんど受けることなく同時に補正することができる。

【００５８】

又、本発明によれば、Ｓ－Ｄプロットにおいて遺伝子発現強度データの和と差を用いることによって、それぞれの蛍光色素の感度の違いが得られ易く、それによる測定誤差を的確に抽出することができるため、蛍光色素の感度の違いによる

測定の歪みを効率良く補正することができる。

【図面の簡単な説明】

【図１】

本発明におけるマイクロアレイの構造を示す図である。

【図２】

本発明の第１の実施の形態の構成を示すブロック図である。

【図３】

本発明の第１の実施の形態の動作を示す流れ図である。

【図４】

本発明の第２の実施の形態の構成を示すブロック図である。

【図５】

チャンネル１で得られたオリジナルデータの遺伝子発現強度の図である。

【図６】

チャンネル２で得られたオリジナルデータの遺伝子発現強度の図である。

【図７】

チャンネル１で得られたオリジナルデータ（第１グリッドから第４グリッド）の遺伝子発現強度の図である。

【図８】

チャンネル２で得られたオリジナルデータ（第１グリッドから第４グリッド）の遺伝子発現強度の図である。

【図９】

オリジナルデータに対するＳ－Ｄプロットである。

【図１０】

チャンネル１のオリジナルデータの遺伝子発現強度の図である。

【図１１】

チャンネル１の第一処理過程後の遺伝子発現強度の図である。

【図１２】

チャンネル１の第二処理過程後の遺伝子発現強度の図である。

【図１３】

チャンネル１の第三処理過程後の遺伝子発現強度の図である。

【図１４】

チャンネル２のオリジナルデータの遺伝子発現強度の図である。

【図１５】

チャンネル２の第一処理過程後の遺伝子発現強度の図である。

【図１６】

チャンネル２の第二処理過程後の遺伝子発現強度の図である。

【図１７】

チャンネル２の第三処理過程後の遺伝子発現強度の図である。

【図１８】

オリジナルデータに対するＳ－Ｄプロットである。

【図１９】

第一処理過程後のＳ－Ｄプロットである。

【図２０】

第二処理過程後のＳ－Ｄプロットである。

【図２１】

第三処理過程後のＳ－Ｄプロットである。

【符号の説明】

１　　入力装置

２　　データ解析装置

３　　出力装置

２１　　データ規準化手段

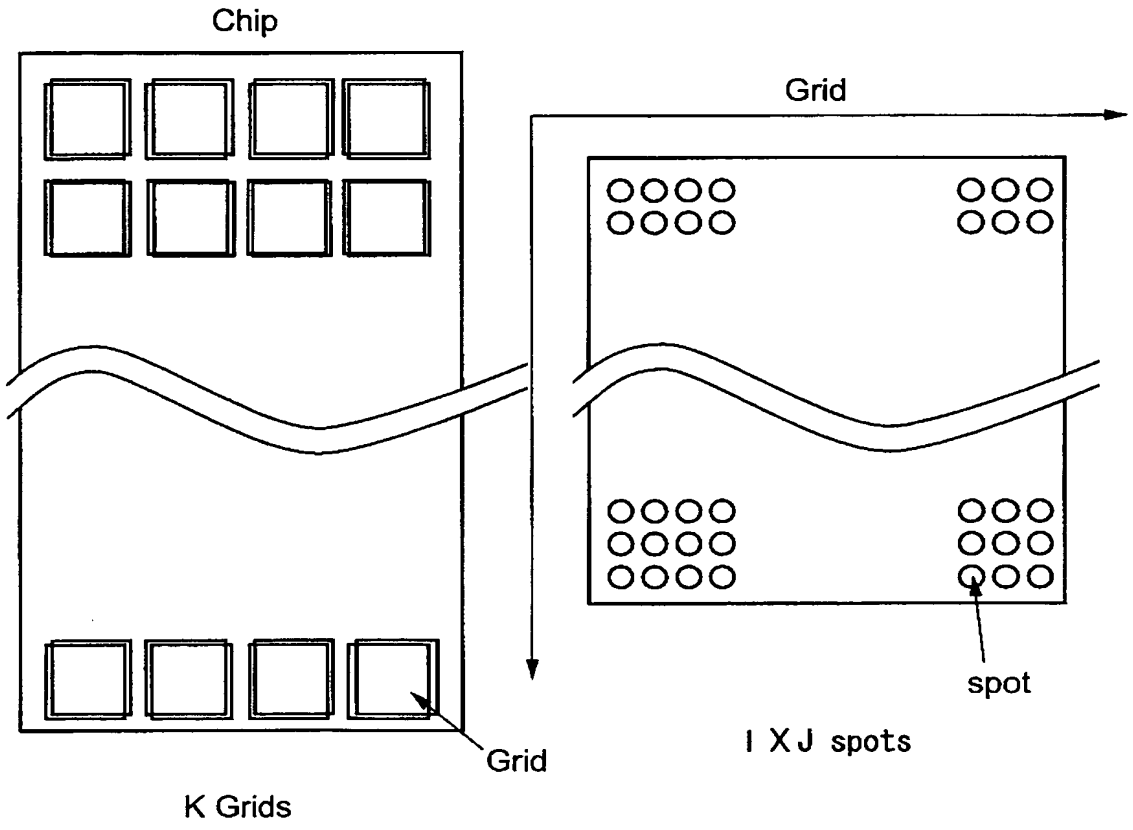
２２　　スポット位置による補正手段

２３　　Ｓ－Ｄプロットによる補正手段

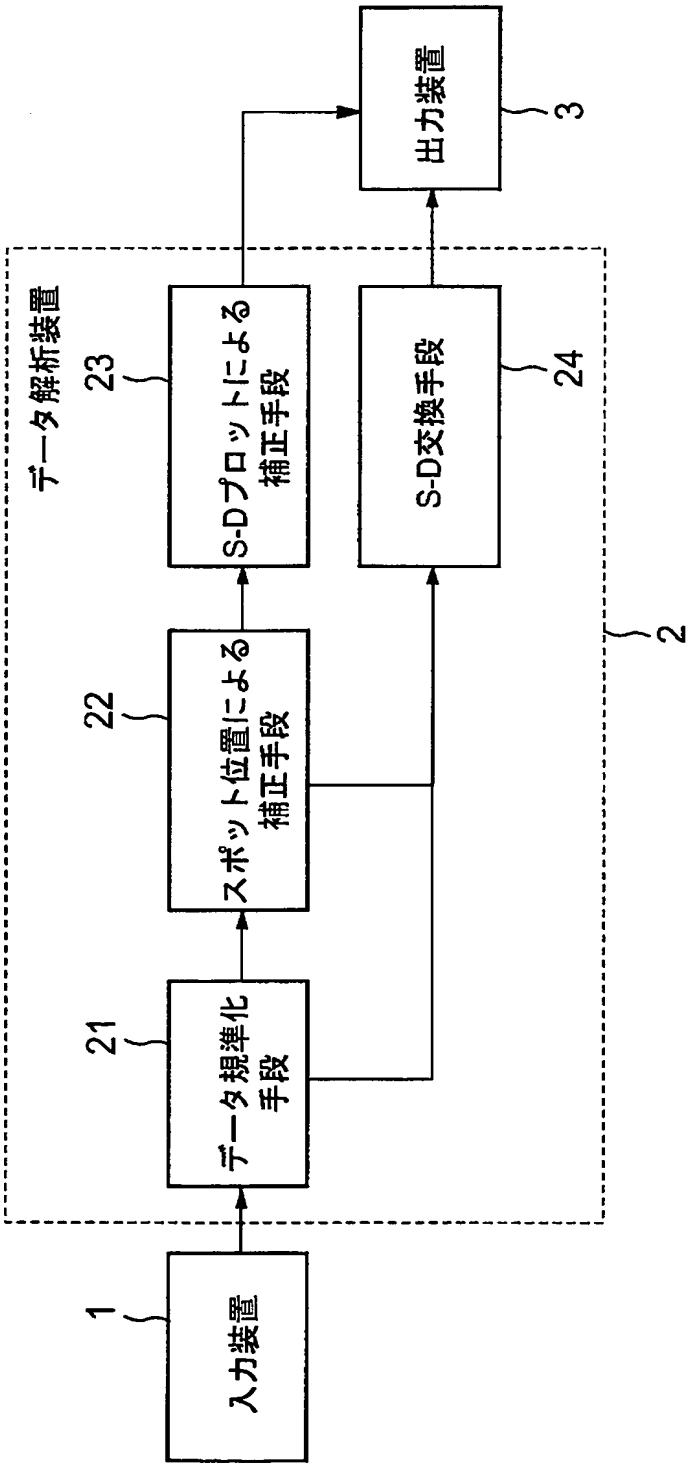
２４　　Ｓ－Ｄ変換手段

【書類名】　図面

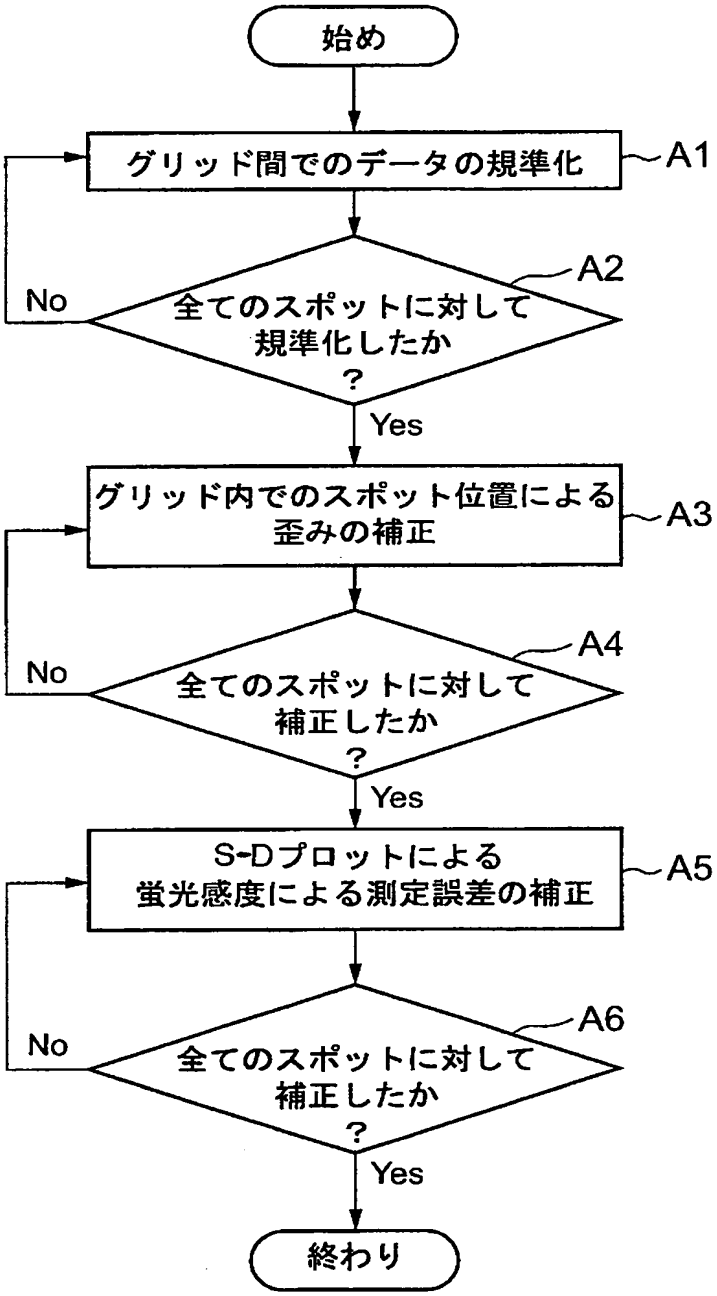
【図１】



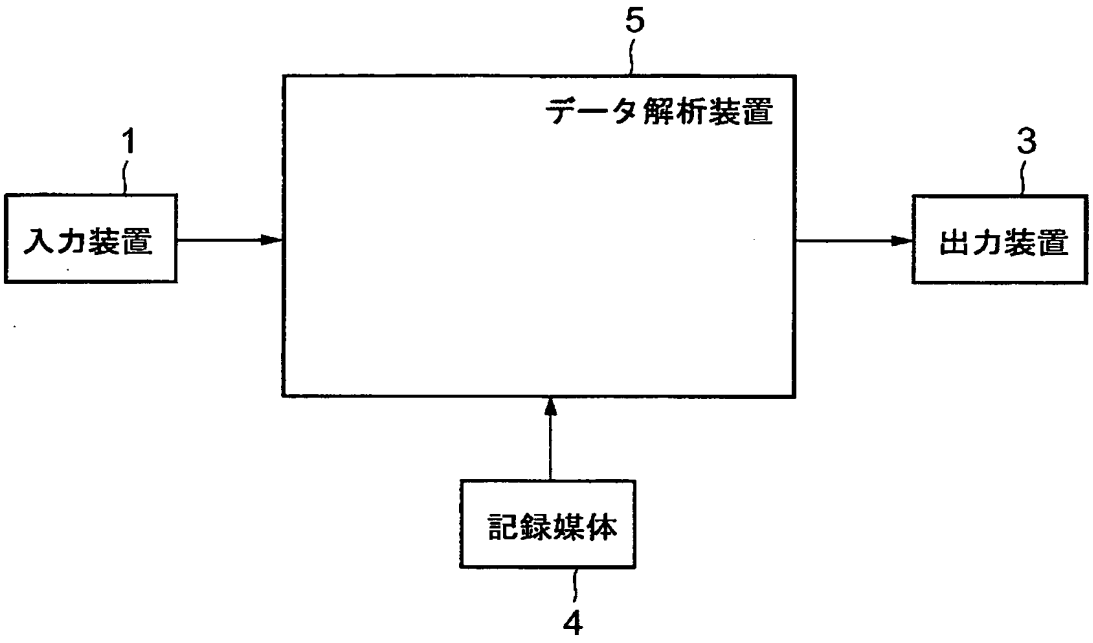
【図２】


出力装置
3
データ解析装置
23
S-Dプロットによる補正手段
S-D交換手段
24
2
22
スポット位置による補正手段
21
データ規準化手段
1
入力装置

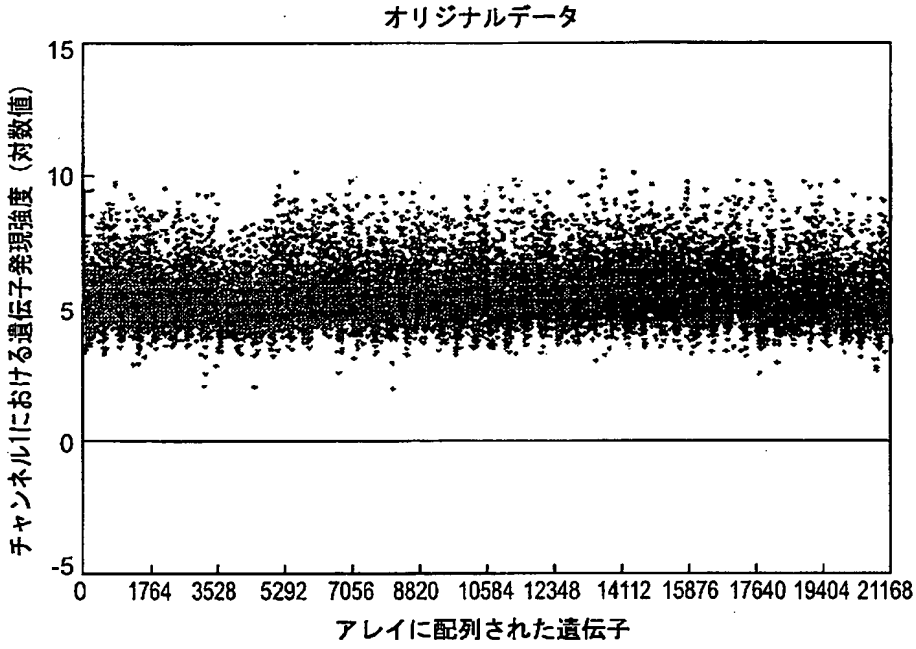
【図３】


始め
グリッド間でのデータの規準化
A1
全てのスポットに対して規準化したか？
A2
No
Yes
グリッド内でのスポット位置による歪みの補正
A3
全てのスポットに対して補正したか？
A4
No
Yes
S-Dプロットによる蛍光感度による測定誤差の補正
A5
全てのスポットに対して補正したか？
A6
No
Yes
終わり

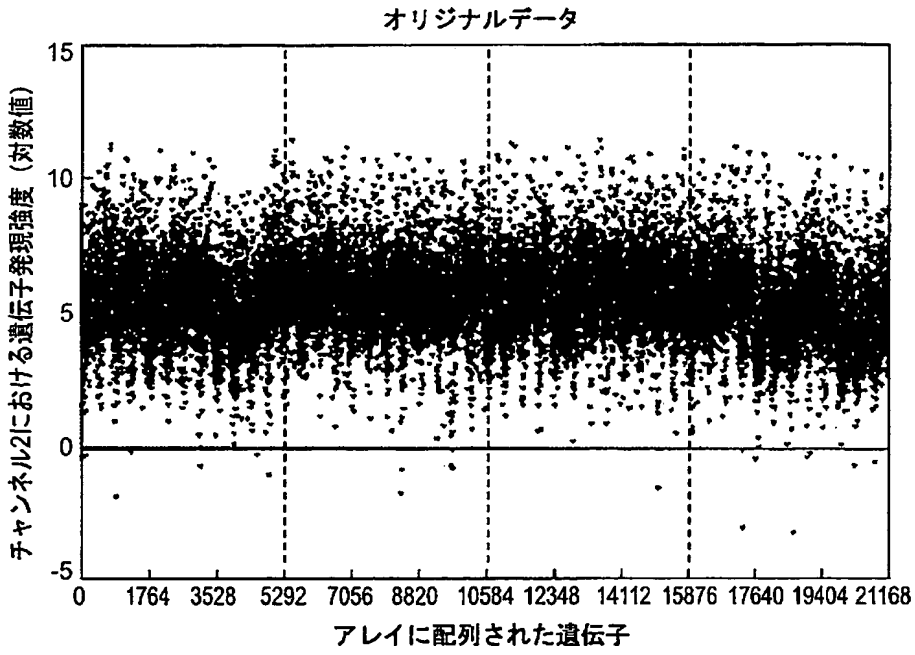
【図４】



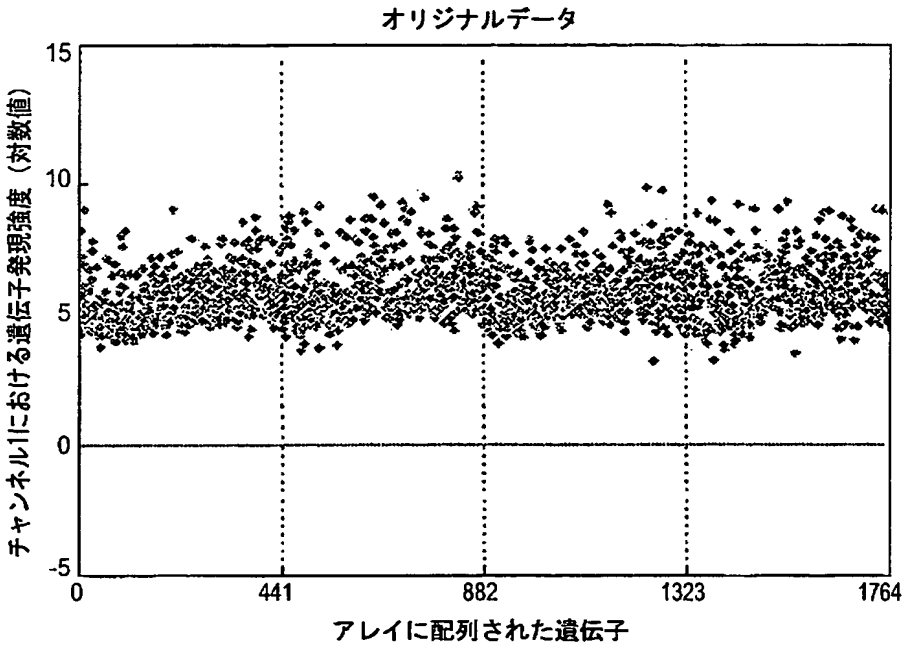
【図５】


オリジナルデータ
チャンネル1における遺伝子発現強度（対数値）
15
10
5
0
-5
0
1764
3528
5292
7056
8820
10584
12348
14112
15876
17640
19404
21168
アレイに配列された遺伝子

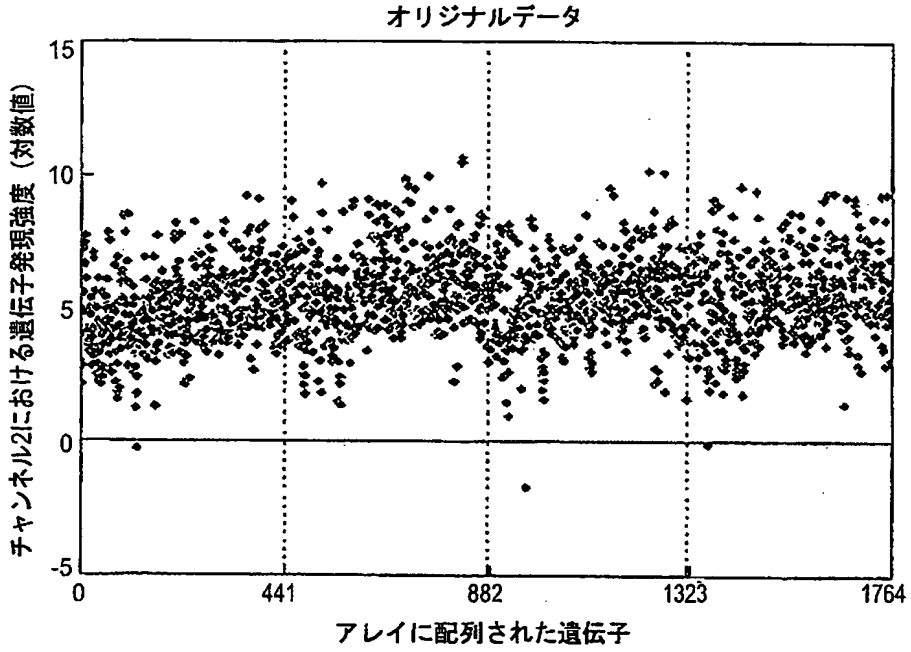
【図６】


オリジナルデータ
チャンネル2における遺伝子発現強度（対数値）
15
10
5
0
-5
0
1764
3528
5292
7056
8820
10584
12348
14112
15876
17640
19404
21168
アレイに配列された遺伝子

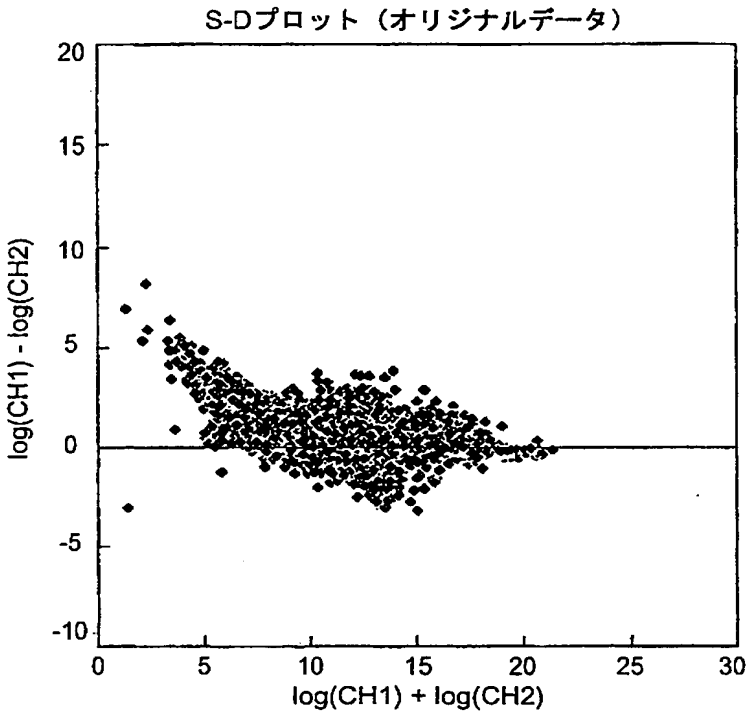
【図７】


オリジナルデータ
チャンネル1における遺伝子発現強度（対数値）
15
10
5
0
-5
0
441
882
1323
1764
アレイに配列された遺伝子

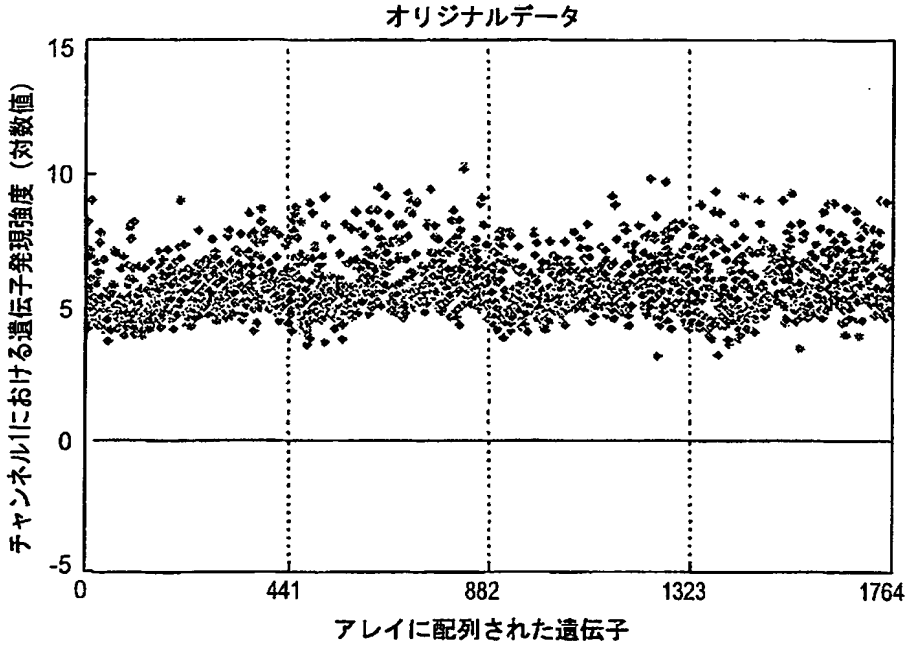
【図８】


オリジナルデータ
チャンネル2における遺伝子発現強度（対数値）
15
10
5
0
-5
0
441
882
1323
1764
アレイに配列された遺伝子

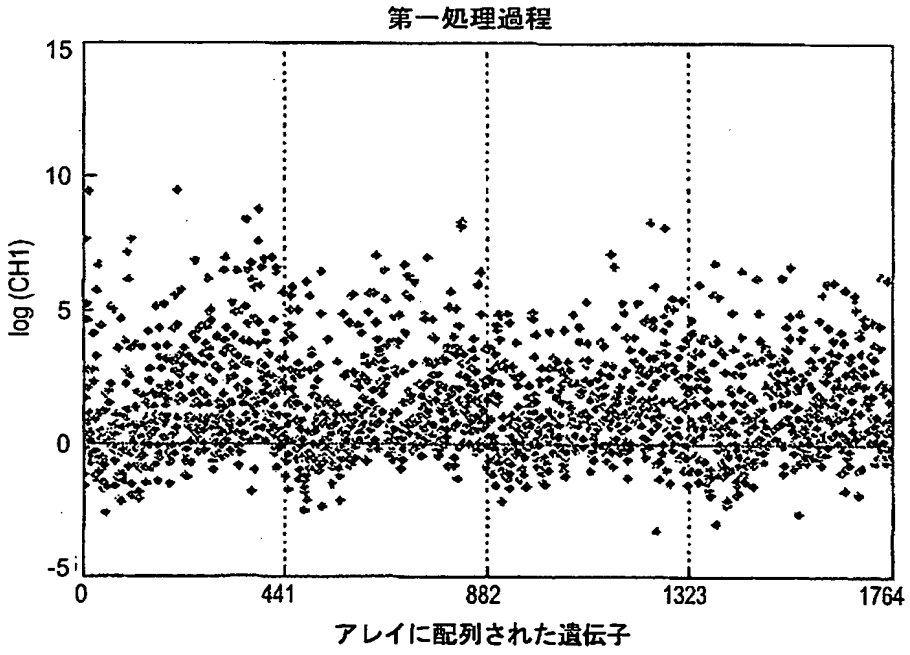
【図９】


S-Dプロット（オリジナルデータ）
log(CH1) - log(CH2)
log(CH1) + log(CH2)
20
15
10
5
0
-5
-10
0
5
10
15
20
25
30

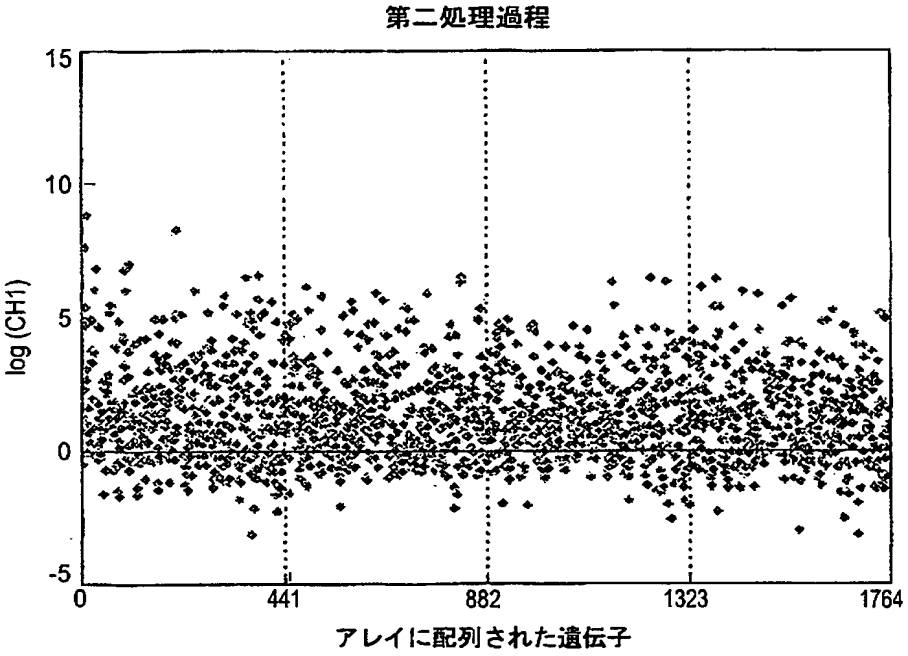
【図１０】


オリジナルデータ
15
10
5
0
-5
チャンネル1における遺伝子発現強度（対数値）
0
441
882
1323
1764
アレイに配列された遺伝子

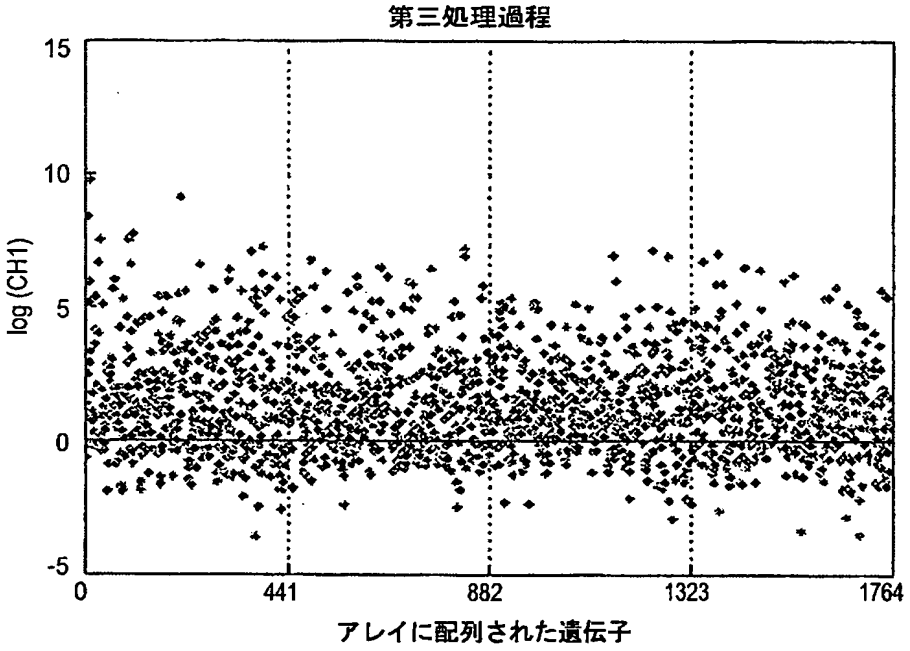
【図１１】



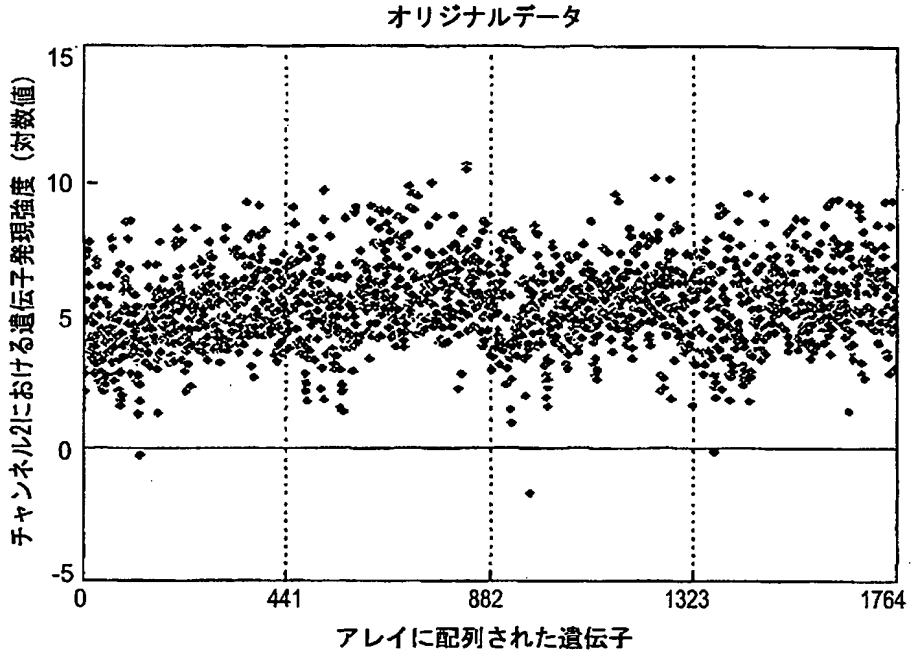
【図１２】



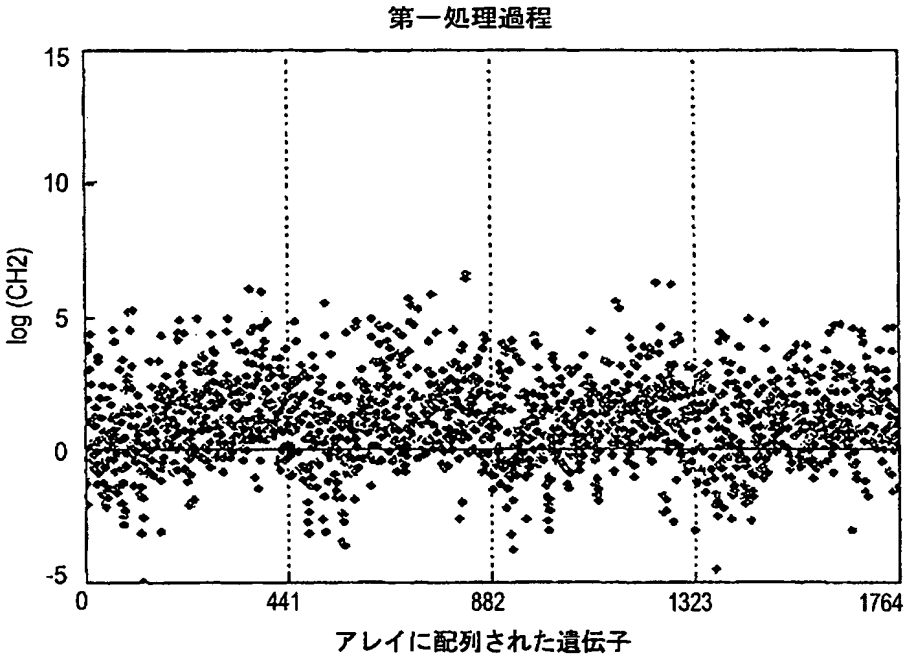
【図１３】



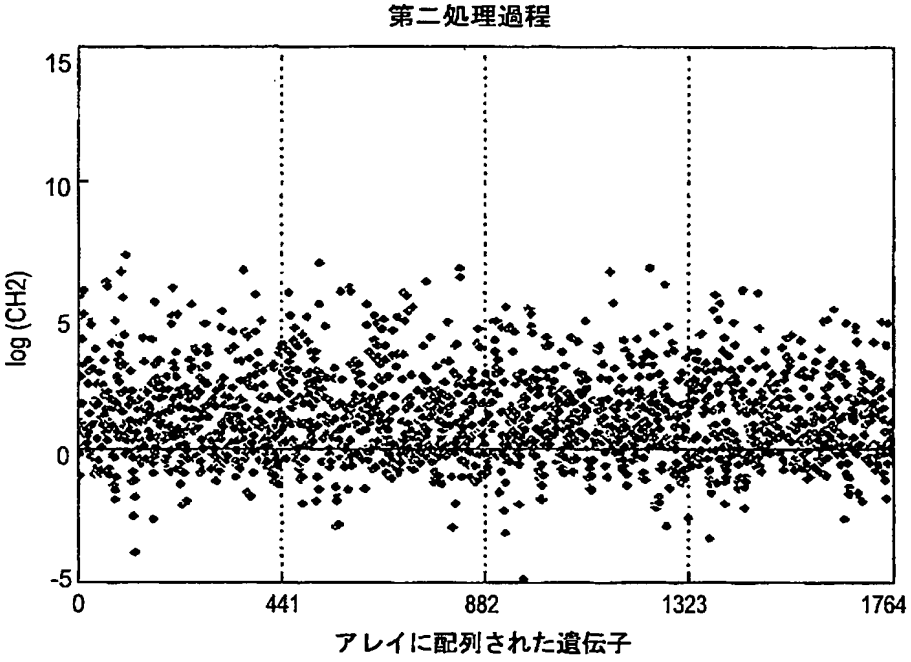
【図１４】


オリジナルデータ
チャンネル2における遺伝子発現強度（対数値）
15
10
5
0
-5
0
441
882
1323
1764
アレイに配列された遺伝子

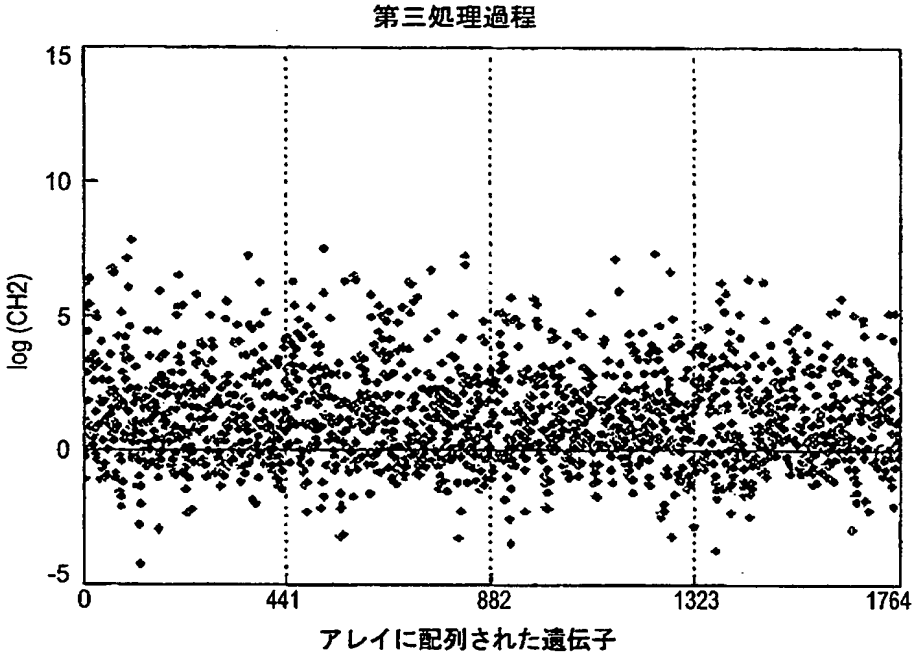
【図１５】



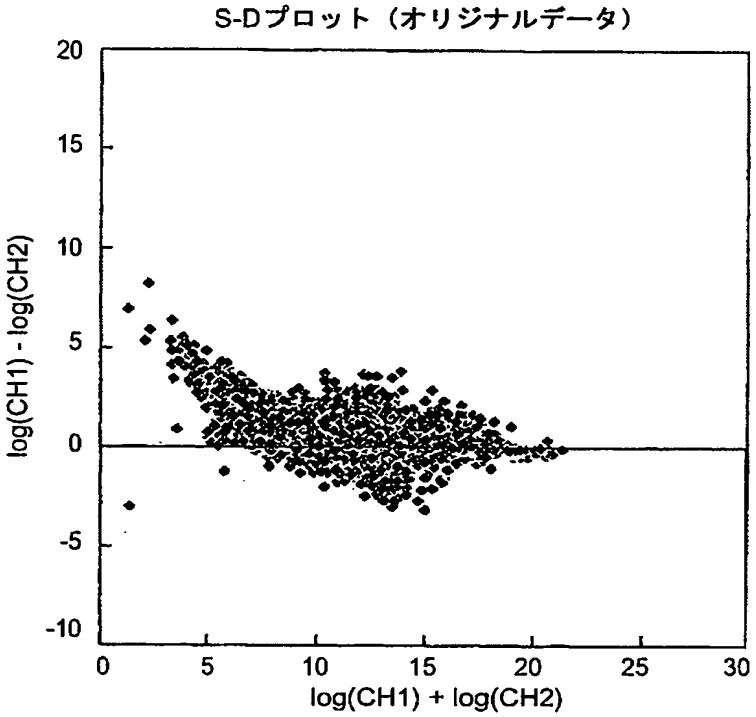
【図１６】



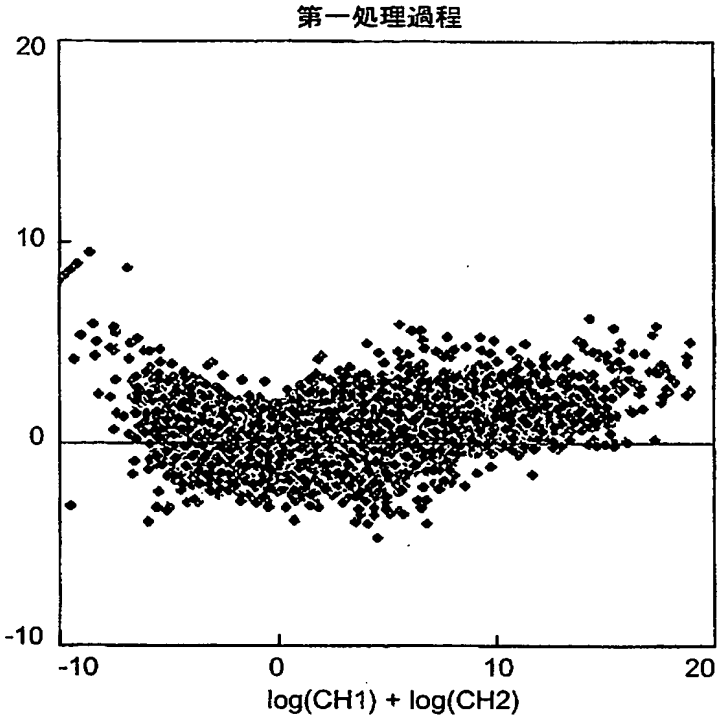
【図１７】



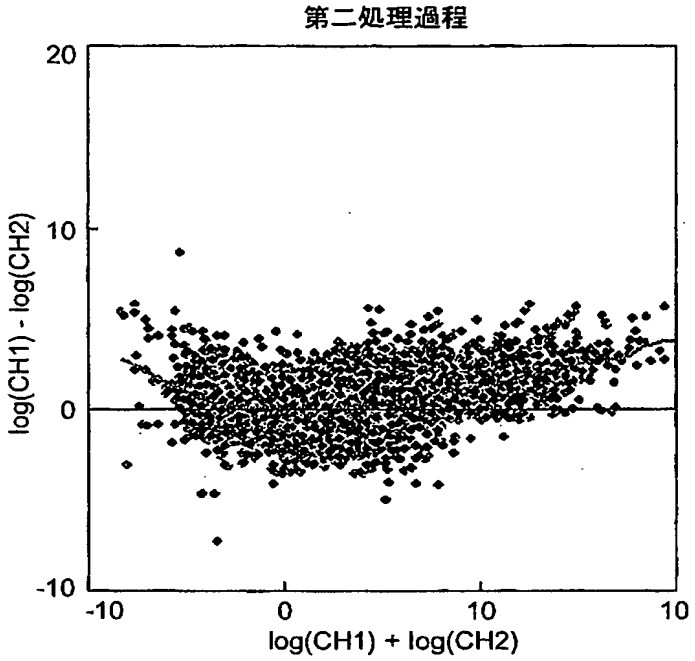
【図１８】


S-Dプロット（オリジナルデータ）
log(CH1) - log(CH2)
log(CH1) + log(CH2)
20
15
10
5
0
-5
-10
0
5
10
15
20
25
30

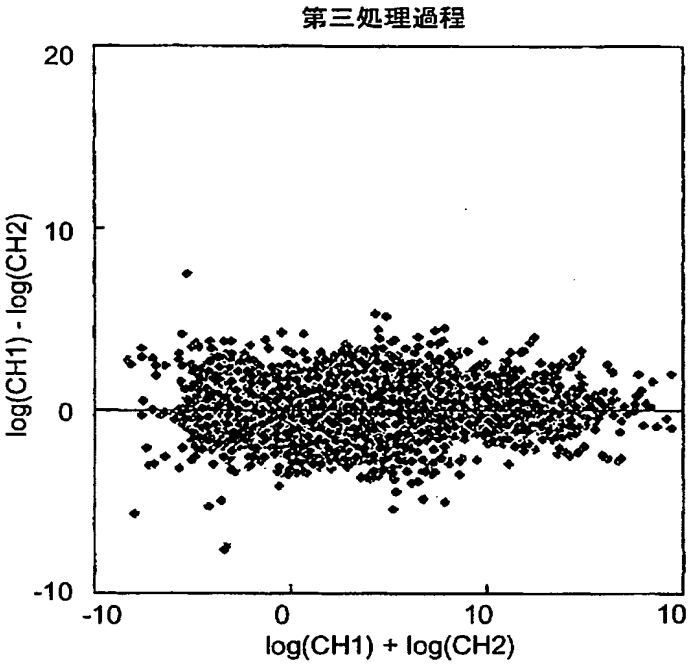
【図１９】



【図２０】



【図２１】



【書類名】　　要約書

【要約】

【課題】　マイクロアレイデータのグローバルおよびローカルな歪みに対してより精密な補正をし、さらに蛍光色素の感度の違いによる測定誤差を補正する。

【解決手段】　第一処理過程であるデータ規準化手段は、入力装置から遺伝子発現強度データを入力し、大半の遺伝子は発現していないことを前提としてグリッド毎の順序統計量を用いて遺伝子発現強度データを規準化し、規準化した遺伝子発現強度データを出力する。第二処理過程であるスポット位置による補正手段は、グリッド毎にスポット位置に依存する歪みをノンパラメトリック平滑化法によって推定し、スポット位置に依存した歪みを補正した遺伝子発現強度データを出力する。第三処理過程であるＳ－Ｄプロットによる補正手段は、Ｓ－Ｄ変換を行い、蛍光色素の感度の違いによる歪みをノンパラメトリック平滑化法によって推定し、蛍光色素の感度の違いによる歪みを補正した遺伝子発現強度データを出力装置に出力する。

【選択図】　　図２

特願２００３－１２４５８５

出　願　人　履　歴　情　報

識別番号　　［０００００４２３７］

１．変更年月日　　１９９０年　８月２９日
［変更理由］　　新規登録
住　所　　東京都港区芝五丁目７番１号
氏　名　　日本電気株式会社

特願２００３－１２４５８５

出　願　人　履　歴　情　報

識別番号　　［５０３０７７１６５］

１．変更年月日　　２００３年　２月２６日
［変更理由］　　新規登録
住　所　　広島県廿日市市宮園９丁目１の７
氏　名　　大瀧　慈

特願２００３－１２４５８５

出　願　人　履　歴　情　報

識別番号　　［５００５３５３０１］

１．変更年月日　　２０００年１１月２０日
［変更理由］　　新規登録
住　所　　東京都中央区八丁堀二丁目２６番９号　グランデビルディング
氏　名　　社団法人バイオ産業情報化コンソーシアム